DOCUMENT RESUME

ED 131 635 EC 091 848

AUTHOR	Fortna, Richard O.; Boston, Bruce O.
TITLE	Testing the Gifted Child: An Interpretation in Lay Language.
INSTITUTION	Council for Exceptional Children, Reston, Va. Information Services and Publications.
SPONS AGENCY	National Inst. of Education (DHEW), Washington, D.C.
PUB DATE	76
NOTE	27p.
AVAILABLE FROM	CEC Information Services and Publications, The Council for Exceptional Children, 1920 Association Drive, Reston, Virginia 22091 ($3.50, Publication No. 142)
EDRS PRICE	MF-$0.83 HC-$2.06 Plus Postage.
DESCRIPTORS	*Gifted; *Screening Tests; *Standardized Tests; *Talented Students; *Talent Identification; Test Construction; Test Interpretation; *Test Reviews

ABSTRACT

Intended for parents and teachers of gifted students (preschool through graduate level), the document provides descriptions in "lay language" of tests used as identification and screening measures in gifted and talented programing. Explained in an introductory section are such concepts as IQ, various uses for tests, types of standardized tests, and kinds of scoring procedures. Sections on what the test measures, content and/or structure of the test, how the test is scored, what the score(s) mean(s), and references for uses of the test in gifted education are provided for each of the following tests: the Wechsler Intelligence Scale for Children-Revised, Stanford Achievement Test Intermediate Level II Battery, California Short-Form Test of Mental Maturity, Tennessee Self Concept Scale--Counseling Form, Torrance Tests of Creative Thinking, and Scales for Rating Behavioral Characteristics of Superior Students. (SBH)

****
* Documents acquired by ERIC include many informal unpublished *
* materials not available from other sources. ERIC makes every effort *
* to obtain the best copy available. Nevertheless, items of marginal *
* reproducibility are often encountered and this affects the quality *
* of the microfiche and hardcopy reproductions ERIC makes available *
* via the ERIC Document Reproduction Service (EDRS). EDRS is not *
* responsible for the quality of the original document. Reproductions *
* supplied by EDRS are the best that can be made from the original. *
****

ED131635

U.S. DEPARTMENT OF HEALTH,
EDUCATION & WELFARE
NATIONAL INSTITUTE OF
EDUCATION

THIS DOCUMENT HAS BEEN REPRODUCED EXACTLY AS RECEIVED FROM THE PERSON OR ORGANIZATION ORIGINATING IT. POINTS OF VIEW OR OPINIONS STATED DO NOT NECESSARILY REPRESENT OFFICIAL NATIONAL INSTITUTE OF EDUCATION POSITION OR POLICY

Testing the Gifted Child:

An Interpretation in Lay Language

by
Richard O. Fortna
Bruce O. Boston

EC091848

ERIC Clearinghouse on Handicapped and Gifted Children
The Council for Exceptional Children

Published in 1976 by
The Council for Exceptional Children
1920 Association Drive
Reston, Virginia 22091

Library of Congress Catalog Card Number 76-46730

Printed in the United States of America

This book is a joint product of the ERIC Clearinghouse on Handicapped and Gifted Children and the ERIC Clearinghouse on Tests, Measurement, and Evaluation. The material about the tests themselves was prepared by Richard O. Fortna of ERIC/TME; Bruce O. Boston of ERIC/HGC provided editorial supervision and wrote the Introduction.

The material in this publication was prepared pursuant to a contract with the National Institute of Education, US Department of Health, Education and Welfare. Contractors undertaking such projects under government sponsorship are encouraged to express freely their judgment in professional and technical matters. Prior to publication, the manuscript was submitted to The Council for Exceptional Children for critical review and determination of professional competence. This publication has met such standards. Points of view or opinions, however, do not necessarily represent the official view or opinions of either The Council for Exceptional Children or the National Institute of Education.

Contents

Introduction	1
Wechsler Intelligence Scale for Children—Revised	5
Stanford Achievement Test: Intermediate Level II Battery	9
California Short-Form Test of Mental Maturity	13
Tennessee Self Concept Scale: Counseling Form	17
Torrance Tests of Creative Thinking	19
Scales for Rating Behavioral Characteristics of Superior Students	23

Introduction

Parents and educators of gifted students, regardless of their feelings and opinions about "testing," continue to be faced with the reality that standardized tests *are* used for identification and will continue to be so used for the foreseeable future. Given this fact, we feel that some basic information needs to be communicated about several of these instruments commonly used as identification and screening measures in gifted and talented programing. Because parents, teachers, school administrators, and educators of the gifted very often lack the expertise to comprehend fully what such tests are and how they function, we have provided descriptions, written in less technical language, which are used for this purpose.

Before proceeding with our main task, however, some comments and disclaimers are in order. First, we hope that by presenting the following descriptions in a form not overburdened by statistical and psychometric language we will be able to demystify them to some degree. Thus, the material about each instrument has been organized in such a way as to answer the questions most often raised by nonspecialists, namely: What The Test Measures, the Content and/or Structure of the Test, How the Test Is Scored, What the Score(s) Mean(s), and some References for Use in Gifted Child Education.

Second, it is our view that the results of no test or test battery can ever totally represent the intelligence or abilities of a given child. Tests are best used as indicators, both of strengths which can be built on and weaknesses which can be remedied. They can also provide clues about programing alternatives. In no case should tests alone be used as screening instruments for gifted and talented programs. They should always be combined with other means of identification, such as teacher nomination, parent interviews and nominations, peer nomination, self appraisals by the child in question, and the opinions of other professionals such as psychologists, physicians, and counselors to form as complete a picture of the child as possible.

Third, we believe it is important to combat some existing misconceptions about the information such instruments actually provide. A good example is the IQ (Intelligence Quotient) test. In the minds of some parents (and often in the minds of some educators who should know better), the IQ obtained for a child, particularly if it serves to label the child as "gifted," achieves an almost mythical quality. It should be pointed out that an IQ is not a figure which stands for a specific amount of intelligence. It is simply a score achieved by a particular child at a particular time on a particular test. It is not a figure forever engraved on Jenny or Johnny's cerebellum and magically extracted by the wonders of modern psychologists and test maker-upers.

At one time IQ's were computed according to a simple mathematical formula. The scores on the IQ subtests were combined to give a total score which represented the child's "mental age." This figure was then divided by the chronological age and multiplied by 100 to get the IQ:

$$\frac{\text{mental age}}{\text{chronological age}} \times 100 = \text{IQ}$$

Thus a 7 year old child with a mental age of 7 would have an IQ of 100; if he had a mental age of 9, the IQ would be 128.

Today, however, the IQ is determined by comparing the child's score with a distribution derived from the scores of all the children who took the test when it was standardized and normed. This average score is given an arbitrary value of 100. IQ scores are then determined by measuring their variance from this mean. Figure 1 may make this clear:

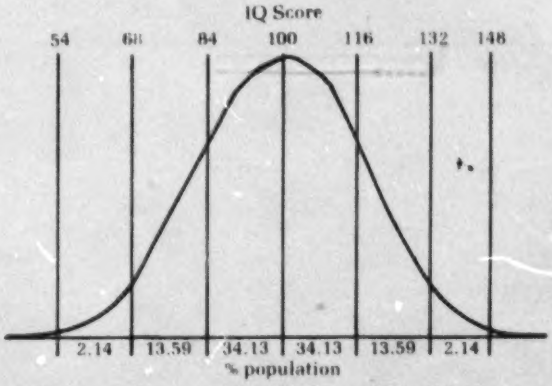

Figure 1. The normal distribution curve for IQ scores.

Thus we see that on a normal distribution curve, 68.26% of the population would have an IQ of between 84 and 116, each within what is called "one standard deviation" from the mean of 100. This does not mean that a child cannot have a lower or higher IQ than 54 or 148 or that these IQ's cannot be measured. It simply means that these children are extremely rare in the population and not appropriately measured by an IQ test.

Parents and teachers of the gifted and talented should realize that the view that the IQ is constant is no longer dominant, nor that IQ is exclusively genetically related, nor that IQ scores rep-

resent nearly all the important cognitive abilities. To the contrary, we now know that IQ scores can and do change, that they are results of both inherited characteristics and learned experiences, that intelligence is multidimensional, and that while IQ tests can measure certain abilities they in no way measure them all. The same holds true for achievement tests.

Schools and school systems have several uses for standardized tests. The material which follows on test use, test types, and scoring has been extracted from a publication by E. Gary Joselyn (*An Introduction to Standardized Testing for Teachers and Administrators*. Minneapolis: University of Minnesota, n.d.).

Standardized tests may be used for *administrative* purposes such as getting a general picture of the level and range of abilities and potential of the students in their charge, for the placement of students in special programs and classes, for making decisions about curriculum, or for evaluating the performances of local children (and hence the schools) by comparing them with national averages. Tests may also be used for purposes of *guidance and counseling* in which the goal is better self understanding on the part of the student. Test scores can be used to help students identify strengths and weaknesses and to make educational, vocational, and career choices. *Instructional* uses by classroom teachers work toward the improvement and individualizing of instruction.

Confusion can arise about different kinds of standardized tests. There are basically four kinds. *Aptitude* tests are designed to measure potential or to predict performance by measuring what the testee can learn. The California Test of Mental Maturity (described below) is this kind of test. *Achievement* tests indicate knowledge of specific kinds of information (e.g., reading comprehension, mathematical competence, knowledge of history, etc.) as well as information deemed "knowable" by the grade group taking the test. *Interest* tests are designed to help students understand their own interests and proclivities and how these may relate to various occupations and careers. Finally, *personality* tests, such as the Tennessee Self Concept Scale described below, include a broad range of instruments which attempt to describe how persons relate to various aspects of their personal and social environment.

Another source of confusion arises from the matter of scoring. There are basically two kinds of scores. *Raw scores* are simply the number of correct responses given to test items, sometimes adjusted for guessing. These have little meaning in and of themselves because different tests and test batteries vary in the number of items included. To give them meaning, raw scores are converted into *derived scores*, of which there are many different kinds. The basic notion of the derived score, however, is that it is a comparison between the score of the child tested and the scores of some known group. This sample to which the test has been given and which provides the frame of reference for evaluating the tested individual is called the *norm group*. The sample is carefully selected in order to assure that it is truly representative of the population to which the test is directed. All scores (except raw scores) are tied to some norm group and are therefore relative and not absolute.

Derived scores on achievement tests (but not IQ tests) fall basically into three categories—grade equivalents, percentile rank, and stanines. *Grade equivalent scores* (GE) are the most common method of reporting performance on achievement tests. They report the average score for students at a particular grade level. If, for example, the average raw score on a test for sixth graders is 50, then 6.0 becomes the GE score for a score of 50. (The 6 stands for the grade, the 0 stands for the month.) While GE scores seem easy to interpret, they are open to misinterpretation. For example:

1. Different tests are normed on different populations. Therefore, students should not be expected to make the same GE score on two different tests which measure a given ability.
2. A GE score does not mean that the materials or instruction appropriate to that level are appropriate for the child tested. A fifth grader who tests at 7.0 for math may or may not be able to handle the material at that level.
3. Because the spread of scores is different on different subtests within a test battery, it is a mistake to assume that identical GE subscores represent equivalent performance as compared with students in the same grade. If a fifth grader scores at 7.0 on both reading and math, it does not mean he has performed equally well in both areas because the spread of scores will be different for each test.
4. Finally, the temptation must be resisted to use GE scores as standards of performance. We often hear people say things like "Twenty percent of our students are below grade level in reading," without realizing that because the GE is an *average* of the scores at a particular grade level, half of any group must be below grade level by definition.

Percentile rank scores (PR) are much less subject to misinterpretation than GE scores. Percentile rank is not the percentage of correct

responses to test items. It is the percentage of students in the norm group who tested at or below the score reported. A PR of 74, for instance, means that the student tested performed as well as or better than 74% of the population on which the test was normed. Because percentile scores tend to bunch between ± one standard deviation from the mean (see Figure 1 above) differences in rank are much more important when they occur at either extreme of the curve than in the middle. Thus the difference between the 90th and the 95th percentile is far greater than the difference between the 48th and the 53rd percentile.

Stanines are a third way of talking about ranking test scores. The name comes from *st*andard scores of *nine* units. Each stanine value represents an approximately equal range of scaled scores. This avoids the problem mentioned above, of overemphasizing differences in the middle of the range which could appear as larger than they are when expressed as percentile ranks.

The important point about scoring procedures, whether by grade equivalent, percentile rank, halves, thirds, quarters, fifths, stanines, or whatever the number of divisions used, is that the number of divisions is totally arbitrary depending on particular uses, preferences, and, likely as not, tradition. There is nothing sacred about them. Moreover, all tests involve some measurement error. The scoring manuals of all reputable instruments report what the "standard error of measurement" is for the test. Thus, if a standard error of measurement is ± 5 points, it means that an IQ test score of 100 has a probability of being between 95 and 105 two-thirds of the time. Test statistics thus represent probabilities, not certainties.

To recapitulate: Tests are instruments for comparing the performance of one group of children against another group of children who took the same test. IQ and achievement tests do not really measure either how "smart" a given child is or what the child has "achieved." They are simply devices for ranking the performance of children against averages which have been established by repeated administration of the same test to large, comparable populations.

The instruments described below were not randomly chosen. ERIC/HGC contacted the state education agency coordinator for gifted and talented programing in states known to have extensive and sophisticated identification procedures to determine what instruments were most frequently used in their programs. Among intelligence tests the Stanford-Binet and the WISC(R) were used about equally; we chose the WISC(R) arbitrarily. The Stanford Achievement Test and the California Test of Mental Maturity were also found to have a high incidence of use; the Tennessee Self Concept Scale was the most widely used personality instrument. The Torrance Test of Creative Thinking is included here not only because of its increasing use as a screening device for gifted and talented programs, but also because it serves to underline a whole different range of abilities from those generally dealt with by intelligence and achievement tests. Finally, we have also included a summary of the Scale for Rating Behavioral Characteristics of Superior Students, developed primarily by Joseph S. Renzulli. While most emphatically not a test, this instrument has achieved widespread use as a preliminary screening device among teachers and program coordinators, and as such, is being encountered by more and more classroom teachers in school systems which are generating or improving programs for the gifted and talented.

Overall, our rationale is simply this. Those concerned with the education of gifted and talented children are entitled to take a look at the instruments used to identify those children, an examination which is not beclouded by the arcane language which too often surrounds them. Freedom of information legislation now assures all parents the right to find out such things as test scores and IQ's, information which in previous times was often issued at the discretion of reluctant school administrators. We hope that the brief descriptions provided to readers here will help them understand more fully what the tests their children take mean.

Before proceeding with the tests themselves, some other points need to be clearly stated. First, anyone interested in a full-blown, technical review of *any* test, including those presented here (except the Renzulli Scale), should refer to the Bible of the testing field, *The Seventh Mental Measurements Yearbook* (O. K. Buros (Ed.). 2 vols. Highland Park NJ: Gryphon Press, 1972). All tests are reviewed in the *Yearbook* by one or more experts in the field. Additional references and bibliography are generously provided.

Second, neither the selection for inclusion nor the description of any of the instruments discussed in the ensuing pages should be construed as an endorsement or recommendation for use of a particular test. The descriptions provided are just that. We have tried to refrain from value judgments and hope that readers will not infer such. What is offered is simply a distillation of information found in test manuals, the tests themselves, and other supporting publications and references.

Third, and finally, readers of the publication

are put on notice that there is a very noisesome and potentially far-reaching wind of controversy blowing across the field of testing. The consensus about testing, once firmly established in American educational circles, is today everywhere under fire. The pervasiveness, profits, and politics of testing are under scrutiny. An entire literature has grown up around the racial, cultural, and socioeconomic biases which have been and are being uncovered. While we do not wish to rehearse the issues raised in that debate here, or to take an editorial position, we would encourage interested readers to consult some of the following resources:

Gartner, A., Greer, C., & Reisman, F. (Eds.). *The new assault on equality: IQ and social stratification*. New York: Harper & Row, 1974.

Gronkind, N. E. *Measurement and evaluation in the classroom (2nd ed.)*. New York: MacMillan, 1971.

Herrnstein, R. J. I.Q. *Atlantic Monthly*, September 1971, pp. 43-64.

Hoffman, B. *The tyranny of testing*. New York: Collier Books, 1964.

Kamin, L. J. *The science and politics of IQ*. New York: John Wiley & Sons, 1974.

McClelland, D. C. Testing for competence rather than for intelligence. *American Psychologist*, 1973, *28*, 1-14.

Mehrens, W. A., & Lehmann, I. J. *Measurement and evaluation in education and psychology*. New York: Holt, Reinhart, & Winston, 1973.

National Elementary Principal, 1975, 54 (4,6).

Perrone, V. (Ed.). *Testing and evaluation*: New views. Washington DC: Association for Childhood Education International, 1975.

Weber, G. *Uses and abuses of standardized testing in the schools*. Occasional papers, No. 22. Washington DC: Council for Basic Education, 1974.

The National Council on Measurement in Education has also produced a series of short monographs (8-10 pages) concerned with the practical implications of educational measurement. These are available at 35 cents each from:

Office of Evaluation Service
Michigan State University
East Lansing MI 48823

Arcesian, P. W., & Madaus, G. F. Criterion-referenced testing in the classroom. Vol. 3, No. 1.

Coffman, W. E. On the reliability ratings of essay examinations. Vol. 3, No. 2.

Cureton, L. W. The history of grading practices. Vol. 2, No. 4.

Ebel, R. L. Shall we get rid of grades? Vol 5, No. 4.

Gardner, E. F. Interpreting achievement profiles—Uses and workings. Vol. 1, No. 2.

Joselyn, E. G., & Merwin, J. C. Using your achievement test score reports. Vol. 3., No. 1.

Mayo, S. T. Mastery learning and mastery testing. Vol. 1, No. 3.

Tyler R. Assessing educational achievement in the affective domain. Vol 4, No. 3.

Warrington, W. G. An item analysis service for teachers. Vol. 3, No. 2.

Wechsler Intelligence Scale for Children—Revised

Wechsler Intelligence Scale for Children—Revised by David Wechsler. Appropriate for use with ages 6 years to 16 years, 11 months. Published by The Psychological Corporation, 304 E. 45th Street. New York NY 10017.

What the Test Measures

This test and earlier versions have been used widely as a measure of intelligence and as a means of appraising learning disabilities since 1949. The author is opposed to the concept of general intelligence as a means of appraising educational, vocational, or other competencies. Instead, Wechsler defines intelligence as "the overall capacity of an individual to understand and cope with the world around him" (Wechsler, 1974, p.5). In order to measure this global capacity, a series of 12 individually administered subtests are administered. Six of these subtests are verbal and six are performance subtests. Ten of the subtests are mandatory and are used to calculate the subject's IQ, while two of the subtests are optional. Each is briefly described in the following section.

Content and/or Structure of the Test

Mandatory Verbal Subtests

Information. This subtest is administered first from the mandatory list of 10. There are 30 questions in this subtest which cover a wide variety of information that children presumably have had an opportunity to acquire. Depending on the age of the subject the examiner begins with questions 1, 5, 7, or 11 and discontinues after five consecutive failures. (Failures are described as a score of zero on an item for all subtests.)

Similarities. This is the third mandatory subtest and includes 17 items which require the subject to explain in what way two things or concepts are alike. All subjects begin with item 1 and the subtest is discontinued after three consecutive failures.

Arithmetic. This subtest is administered fifth. There are 18 problems in this subtest. All but the last three are read to the subject by the examiner and are to be solved without the use of paper and pencil. The last three items are read aloud by the subject. (They are printed on cards and bound into a booklet.) Subjects of different ages start with different items. If the first two items are answered correctly, credit is given for those items skipped. The subtest is discontinued after three consecutive failures.

Vocabulary. This subtest is administered seventh and consists of 32 words presented orally by the examiner. The subject is asked to define the words or "tell what they mean." Subjects of different ages start at different points in the subtest, and if they answer the first two items perfectly, they are given credit for the earlier items. This subtest is discontinued after five consecutive failures.

Comprehension. The ninth subtest consists of 17 items. Each item is designed to measure practical judgment and common sense. Subjects are asked to explain why certain practices are or should be followed or what should be done under certain circumstances. All subjects start with item 1 and the subtest is discontinued after four consecutive failures.

Mandatory Performance Subtests

Picture Completion. This subtest is administered second. This subtest consists of 26 cards, each containing a picture from which some part is missing. The subject is to identify the part that is missing from each picture. Subjects aged 8-16 begin with card 5 and get credit for cards 1-4 if they obtain perfect scores on cards 5 and 6. If either of these cards is failed they then begin with card 1. Subjects ages 6 and 7 begin with card 1. This subtest is discontinued after four consecutive failures.

Picture Arrangement. This subtest is administered fourth. The exercise consists of 13 sets of cards printed with pictures. Each set is to be arranged in an order that tells a sensible story. The first set is a sample item which the administrator puts in proper order to demonstrate what is required of the subject. Subjects aged 6 and 7 begin with item 1, while those 8-16 years of age begin with item 3. If the latter group gets item 3 correct, credit for items 1 and 2 is given. The subtest is discontinued after three consecutive failures.

Block Design. This subtest is administered sixth and it consists of nine blocks or cubes colored red on two sides, white on two sides, and red/white on the remaining two sides. Subjects are to reproduce designs of increasing complexity using from

five to all nine of the cubes. Subjects work directly from block models constructed by the examiner for designs 1 and 2, and from printed cards for designs 3 through 11. The subtest is discontinued after two consecutive failures.

Object Assembly. This is the eighth subtest. The task here is much like a jigsaw puzzle, except that when put together properly each makes a complete object. Each object to be reconstructed is a common, familiar item. The subtest consists of one sample object and four test objects. The entire subtest is given to all subjects.

Coding. The final mandatory subtest requires subjects to associate numbers or shapes with specific symbols. Subjects under 8 are asked to mark 45 shapes with the corresponding mark. Older subjects must code a series of 93 numbers. This subtest is discontinued after 120 seconds. Mazes (described below) may be substituted for the Coding subtest.

Optional Subtests

Digit Span. This subtest consists of two parts: digits forward and digts backward. There are two trials for each of seven test items. If the subject fails on the first trial, a different set of digits is presented. For digits forward, subjects are to repeat from three to nine digits. For digits backwards, subjects are to repeat from two to eight digits. All items are presented orally to the subject and the subject is to repeat orally. The subtest is discontinued after failure on both trials of any item.

Mazes. Nine mazes of increasing difficulty are to be solved by the subject. Subjects of different ages begin with different mazes. Older subjects who obtain a perfect score on the first maze are given full credit for the earlier mazes. The subtest is discontinued after two consecutive failures.

How the Test Is Scored

Scoring is a multistage process, most of it being done while administering the test since the examiner must know when to discontinue a subtest (that is, score one or more responses 0). First, each item of each subtest is scored according to the limits for each item. For example, Information and Picture Completion are scored 1 or 0; Comprehension is scored 2, 1, or 0; Object Assembly is scored 9, 8, . . . , 1, or 0 depending on the time the subject took to complete the task.

Once each item is scored according to the Manual, each subtest total is obtained and entered in the appropriate place under Raw Score on the front of the Record Form. These Raw Scores are then translated into Scaled Scores using the appropriate tables in the Manual—the appropriate table being selected on the basis of the age of the subject. Scaled Scores are entered in the appropriate place for the subtest in question and these Scaled Scores are then totaled to provide Total Scaled Scores for the Verbal and Performance Tests.

Finally, after the total Scaled Scores have been obtained for Verbal and Performance Tests, these scores are summed to obtain a Full Scale Score. These three Scaled Scores are then converted to IQ equivalents using the appropriate tables in the Manual.

Information. The Manual lists one or more acceptable responses for each item. One point is credited for each acceptable response with the maximum raw score being 30 points.

Similarities. Items 1-4 are scored 1 or 0; items 5-17 are scored 2, 1, or 0. The Manual contains scoring criteria and sample responses for each item. Maximum score is 30 points.

Arithmetic. One point is given for each correct response. The Manual lists the correct responses for each item. Maximum score for this subtest is 18.

Vocabulary. Each item is scored 2, 1, or 0. The Manual contains scoring criteria and sample responses for each item. Maximum score for this subtest is 64.

Comprehension. Each item is scored 2, 1, or 0 using the scoring criteria and sample responses in the Manual as a guide. Maximum score is 3[illegible] points.

Picture Completion. Each item is scored 1 or 0 and correct responses are given in the Manual. Maximum score is 26.

Picture Arrangement. Items 1-4 are scored 2 points for passing on the first trial and 1 point for passing on the second trial. The time limit for each trial is 45 seconds. Items 5-12 are scored 3 points for giving the correct arrangement within the time limit, plus a maximum of 2 bonus points per item for quick "perfect" performance based on a table given in the Manual. For items 9-12, there are alternate arrangements that are given partial credit of 2 points, but these are not eligible for time bonuses. Maximum score is 48 points.

Block Design. Designs 1-3 are scored 2 points for passing on the first trial and 1 point for passing on the second trial. Designs 4-11 are scored 4 points for successful completion within the time limit, plus a maximum of 3 bonus points per item for quick "perfect" performance. No credit is

given for partial or incomplete designs. Maximum score is 62 points.

Object Assembly. The scores for these four items depend on the number of cuts correctly joined plus a variable bonus for quick "perfect" performance. Maximum score is 33 points.

Coding. For subjects under 8 years of age, 1 point is scored for each item correctly filled in. For those who have received a perfect score of 45, additional points are given based on the time it took to complete the task. Maximum score is 50 points. For subjects 8 years and older, 1 point is credited for each item filled in correctly. No bonuses are given and the maximum score is 93 points.

Digit Span. Each item is scored 2, 1, or 0. Two points are scored if the subject passes both trials of an item, 1 point is given if the subject passes only one trial, and 0 if neither trial is passed. Maximum score for this subtest is 28.

Mazes. Each maze is scored on a multipoint basis from full credit when the subject solves the maze within the time limit without making any errors to no credit. No credit is determined in two ways, first by the number of errors committed and second when the maze is not solved within the time limit. (An error is defined in this task as an entrance into a blind alley, that is, when the subject makes a clear crossing of an imaginary line across the mouth of the blind alley.) Maximum score for this subtest is 30 points.

What the Score(s) Mean(s)

A variety of scores are obtained from the WISC(R): Raw Scores, Scaled Scores, and Intelligence Quotients (IQ's).

The number of scores is almost mind boggling—twelve different Raw Scores (one for each subtest), 15 different Scaled Scores (one for each subtest plus a total for the Verbal, Performance, and Full Scale), and 3 different IQ's (Verbal, Performance, and Full Scale). That is a total of 30 scores!

Depending on the circumstance and the purpose, each variety and type of score obtained from the WISC(R) mentioned above could be interpreted. However, we are limiting our discussion to the IQ scores obtained. Readers who wish to learn more about "psychological diagnosis" or "individual assessment" are urged to refer to the references listed in Buros (1972).

An IQ of 100 represents the performance of the average child of a given age. Approximately two-thirds of all children obtain scores between 85 and 115, while 95% score in the range of 70 to 130. The following verbal descriptions are given to selected IQ ranges:

Very superior	130 & Above
Superior	120-129
High average or bright	110-119
Average	90-109

In most instances scores are reported separately for the Verbal and Performance Tests in addition to the Full Scale IQ. The rationale for reporting scores on the Verbal and Performance scales separately is based on the research supported contention that these are the two principal modes by which human abilities are expressed. Further, trained psychologists, psychometrists, and counselors use differences or similarities between the two scores to diagnose specific problems or to indicate that additional testing is needed.

Caution is urged when attempting to interpret differences between the Verbal and Performance IQ's. Even though each IQ is quite reliable,* the reliability of the difference between them is lower. Thus, it is generally accepted that the two scores must differ by at least 15 IQ points before the difference can be taken seriously.

References for Uses of the Test in Gifted Education

Glasser, A. J., & Zimmerman, I. L. *Clinical interpretation of the Wechsler Intelligence Scale for Children (WISC)*. New York: Grune & Stratton, 1967.

Littell, W. M. The Wechsler Intelligence Scale for Children: Review of a decade of research. *Psychological Bulletin*, 1960, 57, 132-156.

Wechsler, D. Cognitive, conative and non-intellective intelligence. *American Psychologist*, 1950, 5, 78-83.

Wechsler, D. *Manual for the Wechsler Intelligence Scale for Children—Revised*. New York: The Psychological Corporation, 1974.

Wechsler, D. *The measurement and appraisal of adult intelligence*. (4th ed.). Baltimore: Williams & Wilkins, 1958.

*Editor's note: In the testing field, "reliability" refers to the degree to which the test consistently measures what it purports to measure.

Stanford Achievement Test Intermediate Level II Battery

Stanford Achievement Test: Intermediate Level II Battery—1973 Revision by Richard Madden, Eric F. Gardner, Herbert C. Rudman, Bjorn Karlsen, and Jack Merwin. Appropriate for use with students in grade levels 5.5 through 6.9. Published by Harcourt Brace Jovanovich, Inc., 757 Third Avenue, New York NY 10017.

Intermediate Level II has been chosen for descriptive purposes since it covers the middle grades. Readers should be aware that, in all, the Stanford Achievement Test Series covers the grade range 1.5 through 12, as well as college. Primary Levels I, II, and III cover the range 1.5 through 4.4. The two Intermediate Levels cover the range 4.5 through 6.9. One Advanced Level is appropriate for use in Grades 7 through 9.5. Two TASK (Test of Academic Skills) Levels range from Grade 8 through 12 while the final TASK level is appropriate for Grade 13. Details about other levels may be found in the Harcourt Brace Jovanovich *Catalog of Tests and Related Services* (1974).

What the Test Measures

There are two "versions" of this battery. The Complete Battery consists of 11 subtests: Vocabulary, Reading Compprehension, Word Study Skills, Mathematics Concepts, Mathematics Computation, Mathematics Applications, Spelling, Language, Social Science, and Listening Comprehension. The Basic Battery omits Social Science, Science, and Listening Comprehension. Working times for the two are 5 hours, 20 minutes and 3 hours, 45 minutes, respectively. Each of the subtests is described in the section that follows.

Content and/or Structure of the Test

Vocabulary. The 50 items and four option choices of this test are dictated by the examiner to the examinees. Examinees also read the item and choices in their test booklets. Vocabulary items are classified into five instructional objectives: (1) reading and literature, (2) nonfiction and reference, (3) mathematics and science, (4) social science, and (5) arts and crafts.

Reading Comprehension. There are 71 items in this test distributed among several paragraphs of widely varied content. All are four option multiple choice items. Instructional objectives for this test are grouped into five areas: (1) global meaning, (2) explicit meaning, (3) implicit meaning, (4) meaning determined by the context, and (5) inferential meaning.

Word Study Skills. This 50 item test has two parts, one covering the skills of phonetic analysis, the other covering structural analysis. Phonetic analysis consists of 25 four option, multiple choice items classified into three instructional objectives: (1) consonant sounds, (2) short and long vowel sounds, and (3) variant vowel sounds. Structural analysis consists of 25 five option, multiple choice items classified into five instructional objectives: (1) affixes, (2) syllables (the pupil blends parts of words which have been divided at a point where there are two consonants between vowel sounds), (3) syllables (words divided where there is one consonant between two vowel sounds), (4) accent, and (5) disconnected syllables (no word can be made from the syllables given).

Mathematics Concepts. This test consists of 35 four option, multiple choice items grouped into four instructional objectives: (1) number, (2) notation, (3) operations, and (4) geometry, measurement, set notation, ratio, function, logical thinking, and concepts not directly involving numbers.

Mathematics Computation. This test has two parts. Part A consists of 24 three option, multiple choice items which require the test taker to choose the proper sign (>, <, or =) to make the mathematical statement for that item true. Part B consists of 21 five option, multiple choice items. Test takers must solve the problem and choose the correct answer, including an option labeled "NH," meaning that the answer to this problem is "not here." These items are classified into four instructional objectives: (1) knowledge of primary facts and solution of simple mathematical sentences, (2) addition and subtraction algorithms, (3) multiplication and division algorithms, and (4) common fractions.

Mathematics Applications. This test consists of 40 "word problems." Each item is a five option, multiple choice item with an option "NH," meaning the answer is "not here." These problems are classified into the following five instructional objectives: (1) selection of an appropriate operation, (2) analysis and development of a solution design, (3) rate and scale problems, (4) measurement, and (5) graph reading and interpretation.

Spelling. Each of the 60 items in this test consists of four options, one of which contains a misspelled word. The first eight items present the four words imbedded in a short phrase and underlined. The remaining items simply present four words. The instructional objectives for this test are grouped into three categories: (1) homophones, (2) phonics, and (3) word building.

Language. This test consists of 80 items and is presented in four distinct parts each of which utilizes a different item type. Part A utilizes a four option, multiple choice format but is essentially a "fill in the blank" type of item. Part A has 42 items. Part B, which has 8 items, requires the test taker to read a group of words and decide whether they represent a complete sentence as printed (even though something could be added to them) or whether the group is not complete. In the latter case the examinee must choose from the other three options the one which could be added before or after in order to make a complete sentence. Part C, 20 three option items, requires the examinee to read a group of words and decide whether the group for that item makes one complete sentence, two or more sentences, or does not make a complete sentence. Part D, 10 four option items, presents dictionary definitions for words at the bottom of the page, then presents items with those words imbedded in different sentences. The first two sentences (items) require the test taker to use the dictionary definition to choose the best answer for the context. The three remaining sentences (items) ask related questions about the word. The items in this test are classified into three broad instructional objectives: (1) conventions, which include capitalization, punctuation, and usage; (2) language sensitivity; and (3) dictionary skills.

Social Science. This 54 item test utilizes five item groupings in order to facilitate the analysis of achievement. Some groupings require the examinee to interpret maps, graphs, charts, written passages, et cetera. All items are four option, multiple choice even though the stimulus material differs. The items are classified into five broad instructional areas: (1) geography, (2) history, (3) economics, (4) political science, and (5) sociology and anthropology.

Science. This is a 60 item, four option, multiple choice test. Six item groupings have been used to facilitate the analysis of achievement. Some groupings are based on pictures, graphs, results of experiments described to the test taker, et cetera. All groupings contain a high degree of process or concept oriented items. Instructional objectives used to classify the items are: (1) concepts of matter, (2) concepts of energy, (3) change in the physical environment, (4) form and function of living things, (5) environmental interaction, and (6) basic skills of science.

Listening Comprehension. This is a 50 item, four option, multiple choice test. Students have no test booklet for this test simply an answer sheet. Items have been grouped into the following instructional objectives: (1) global meaning, (2) explicit detail, (3) implicit meaning, (4) conceptual meaning and classification, and (5) inference and logical analysis.

How the Test Is Scored

Basically, there are two methods of scoring, by hand or by machine. Each form of the tests has its own hand-scoring keys or stencils. There are 17 stencils for each form; therefore it is imperative that the appropriate stencil for the particular form administered be used. Directions, both general and specific, for hand scoring are with the keys.

The score for each test is the number of correct responses. The scorer, following the directions for the specific test in question, simply counts the number of correct responses and enters that number in the appropriate place on the answer sheet or test booklet. Using the Manual (Madden, et al., 1974 b), the number right or total number of correct responses is translated into a scaled score, grade equivalent, percentile rank, or stanine.

Inasmuch as there are 595 items in the Complete Battery, it may not be advisable to plan to hand score many Complete Batteries. Whatever is hand scored should be checked, preferably by a second person, to insure as great an accuracy as possible.

Many forms of machine scoring are available for this Battery. The publisher offers some special guidelines which should be observed in preparing for machine scoring:

1. Make sure that the name grid is filled out correctly.
2. Make certain that all necessary information beyond the name is correctly coded.
3. Check to be sure all circles have been fully filled in, using a soft lead pencil. No crayons.
4. Erase any stray marks on answer documents.
5. Organize and ship answer documents according to the instructions provided with MRC materials (Madden, et al., 1974 b).

A wide variety of machine produced reporting forms are available for this Test Battery. Details may be found in the Harcourt Brace Jovanovich *Catalog of Tests* (1974).

What the Score(s) Mean(s)

This section describes the three types of scores used in reporting results for the Stanford Achievement Test Battery—grade equivalents, percentile ranks, and stanines.

Individuals who interpret achievement test scores like those of school children should do so in light of the pupils' school experience. There are a number of other factors which affect school achievement in addition to the school curriculum and the teaching. All scores can be interpreted more precisely and fairly when other factors such as the following (taken from Madden, et al., 1974 c, p. 13) are given proper consideration as factors which may affect achievement both positively and negatively:

1. The grade level of the pupil.
2. The number of years the pupil has attended school.
3. Age.
4. Out of school experiences.
5. Health and nutrition.
6. The personal attributes of a child, such as emotional adjustment, self image, et cetera.
7. Socioeconomic status.
8. The immediate quality of the school environment.
9. Any bias in the test content.
10. The testing situation.

A detailed discussion of the meaning of scores on the various tests of this battery may be found in the Teacher's Guide for Interpreting (Madden, et al., 1974 c). In addition to a discussion of how the score for a particular test might be interpreted, the Manual discusses interrelationships among the other tests in the battery and provides suggestions for teaching the various subjects or skills.

Norms for the Stanford Achievement Test were obtained from testing representative national samples of children twice (fall and spring) during each school year and three times in Grades 1 and 2. These norms provide information about how pupils from some defined group typically perform on a given test, *not* how they ought to perform. While five types of norms are presented for the Stanford, only three are mentioned here.

Grade Norms and Grade Equivalents

When standardized tests were first introduced, many manuals presented Grade Norms and translated raw scores into "grade equivalents." These equivalents were somewhat analogous to the concept of mental age. A "grade equivalent" of 6.0 is assigned to the score which the average beginning sixth grader makes.

Norms are not standards. It is a common mistake to assume that all pupils in the sixth grade should reach the sixth grade norm. That is, if tested in the second month of the sixth grade, everyone should be at a grade equivalent of 6.2. This, of course, is a fallacy—fifty percent of the students in the standardization sample fall below the norm. Grade norms also tend to imply that two pupils with a grade equivalent of 7.0 are similar, even if one is in Grade 3 and one is in Grade 9. Grade norms can lead teachers, students, and parents to unsound conclusions; therefore, these scores should be interpreted not in isolation but along with the other two scores provided by the battery—percentile ranks and stanines.

Percentile Ranks

In simplest terms these scores are ranks expressed in percentage form. A person's percentile score tells what proportion of the group falls below that point. For example, take a sixth grade pupil with a percentile rank of 66 on the Mathematics Computation test. This score indicates that this pupil has scored higher on this test than 66% of the group in question. The group in question means, "What reference group are you using to obtain the percentile rank?" This pupil's score could have been compared to: "National Norms," "Local Norms," "National Norms for Eighth Grade Pupils," et cetera.

Stanines

This word is derived from standard scores of nine units. Each stanine value represents approximately equal ranges of scaled scores, i.e., one-ninth of the total range. This helps to lessen the problem of overemphasizing small, insignificant differences in the middle range. As a general rule, stanine interpretation states that if a pupil's score varies by as much as two stanines from one test to another, this probably represents a "true" difference in performance. As with percentiles, it is important to know on which group the stanines are based.

References for Uses of the Tests in Gifted Education

Harcourt Brace Jovanovich. *Catalog of tests and related services.* New York: Harcourt Brace Jovanovich, 1974.

Madden, R., et al. *Stanford Achievement Test: Intermediate Level II Battery. Manual, Part I, Teacher's directions for administering.* New York: Harcourt Brace Jovanovich, 1974(a).

Madden, R., et al. *Stanford Achievement Test: Intermediate Level II Battery. Manual, Part II. Norms booklet.* New York: Harcourt Brace Jovanovich, 1974(b).

Madden, R., et al. *Stanford Achievement Test: Intermediate Level II Battery. Manual, Part III. Teacher's guide for interpreting.* New York: Harcourt Brace Jovanovich, 1974(c).

Madden, R., et al. *Stanford Achievement Test: Intermediate Level II Battery. Manual, Part IV. Administrator's guide.* New York: Harcourt Brace Jovanovich, 1974(d).

Madden, R., et al. *Stanford Achievement Test: Intermediate Level II Battery. Manual, Part V. Technical data report.* New York: Harcourt Brace Jovanovich, 1974(e).

California Short-Form Test of Mental Maturity

California Short-Form Test of Mental Maturity: Level 2H by Elizabeth T. Sullivan, Willis W. Clark, and Ernest W. Tiegs. Appropriate for use with students in Grades 6 and 7. Published by CTB/McGraw-Hill, Del Monte Research Park, Monterey CA 93940.

Level 2H has been chosen for descriptive purposes since it covers the middle grades. Readers should be aware that, in all, the California Short-Form Test of Mental Maturity covers the grade range Kindergarten through College, and Adults. Level 0 is for use with Kindergarten and entering Grade 1 students; Levels 1 and 1H cover Grades 1 through 4; Levels 2 and 2H cover Grades 4 through 7; Level 3 should be used in Grades 7 and 8; Level 4 is appropriate for use in Grades 9 through 12; and Level 5 may be used in Grade 12, College, and with Adults. Details about the other levels may be found in the CTB/McGraw-Hill *Catalog* (1974).

What the Test Measures

This instrument, "like the parent California Test of Mental Maturity, provides information about the functional capacities that are basic to learning, problem-solving, and responding to new situations" (California Test Bureau, 1963). The test consists of seven administrative test units, each of which measures some aspect of mental ability and contributes to a pattern of summary and derived scores. At each of the levels of the test, the rate and scope of mental development are measured in terms of four statistically derived factors: Logical Reasoning, Numerical Reasoning, Verbal Concepts, and Memory. Within these factors, the seven test units are grouped into two sections: Language and Non-Language. Each of the seven test units are described in more detail in the next section under the factor to which they belong.

Content and/or Structure of the Test

Factor I—Logical Reasoning

Test 1: Opposites. This test unit consists of 15 items requiring 4 minutes testing time. Each item consists of a row with five pictures. The first picture in the row is the stimulus and examinees are to choose the one from the remaining four that shows something opposite.

Test 2: Similarities. This unit has 15 items, each consisting of seven drawings in two groups. The three objects shown in the first group are alike in some respect. The examinee is to determine what attribute the three have in common, then find something in the second group which is logically related to the first group.

Test 3: Analogies. This test unit also has 15 items. Each item has seven drawings in two groups. The first three set up the analogic relationship which is to be determined by the examinee who must then find one of the four remaining pictures which relates to the third picture in the same way as the first and second pictures relate. In verbal terms the items read: The first picture is to the second picture as the third picture is to which of the remaining four pictures?

Factor II—Numerical Reasoning

Test 4: Numerical Values. This unit has 15 items based on manipulating combinations of coins. The examinee is given information about money (coins from 1 cent to a half-dollar in the form, "One quarter is 25 cents, 1 dime is 10 cents, etc."). Nineteen answers, representing different combinations of the five coins, are also presented. The examinee is to determine which of the four options (taken from the 19 answers) given for each item will solve the problem of "X coins equal Y cents."

Test 5: Number Problems. These 10 word problems present a numerical situation and ask a question about it. Examinees are to solve the problem, mentally or with scratch paper if necessary, and select the correct answer from the four options presented.

Factor III—Verbal Concepts

Test 6: Verbal Comprehension. This unit consists of 25 vocabulary items requiring the examinee to select a synonym for the word presented from the four options available.

Factor IV—Memory

Test 7: Delayed Recall. The 25 items in this unit test are based on a story read to the examinees at the beginning of the test period, that is, before any of the first six tests are administered. Each item relates to facts or ideas stated or clearly

implied in the story. Examinees have to choose the correct response from the four options presented.

How the Test Is Scored

There are basically two methods of scoring, by hand or by machine. Responses marked directly in the test booklets can be hand scored only. Most of the machine scorable answer sheets, however, may be hand scored.

The Examiner's Manual (p. 25) lists six standard rules for scoring which are followed by all California Test Bureau (CTB) scoring centers. These rules should be followed by all who score the Short-Form.

1. The score for each test is the number of right responses.
2. An item for which two or more responses are marked is not counted in the scoring.
3. Items omitted are not counted in the scoring.
4. If three or more items on the complete Short-Form have more than one mark, the test results are considered void on the grounds that the pupil did not understand or did not follow directions.
5. Nonattempted test units are considered void. If no item in a separately-timed unit of the Short-Form has been marked, that test unit is declared nonattempted. If one or more items in a timed unit have been marked—even if all responses are wrong—that unit has been attempted, and the score may be combined with other acceptable test unit scores to obtain part scores.
6. No derived score may be obtained for the total, the section (Language or Non-Language), or for the factor containing a voided test unit.

A variety of machine-scoring answer sheets and answer cards is available to users. Specific instructions for those offered by CTB scoring centers is described in the Examiner's Manual (pp. 25-27). Readers wishing more information on any of these answer forms are referred to the Manual.

What the Score(s) Mean(s)

Five types of scores are obtained from the Short-Form: raw scores, standard scores, percentile scores, stanines, and intelligence quotients.

Raw scores are obtained by counting the number of correct responses for each of the test units. Raw scores for Tests 1-3 are summed to get the total raw score for Logical Reasoning, raw scores for Tests 4 and 5 are summed to get the total raw score for Numerical Reasoning, and Tests 6 and 7 represent the total raw scores for Verbal Concepts and Memory, respectively. By combining the raw scores for Tests 1, 2, 3, and 4, the raw score for the Non-Language Section is obtained. The raw score for the Language Section is the total of Tests 5, 6, and 7. Total test raw score then is the sum of the two just obtained.

Standard scores, percentile scores, and stanine equivalents for raw scores for the four factors, the Language and Non-Language Sections, and the Total are obtained using tables found in the Examiner's Manual. IQ's for the Language, Non-Language, and Total scores are also obtained by means of tables in the Examiner's Manual. The "deviation IQ" adopted in the 1963 Revision is designed to provide a constant mean (average) of 100 and a standard deviation of 16 IQ points for all age levels (p. 34).

The publisher does not necessarily recommend one score over the other. In fact, one gets the distinct impression that by stressing the use of the individual profile, the publisher feels the best information about an individual's performance is obtained when all of the scores are used in concert with one another.

After plotting scores on the individual profile, the Manual (p. 27) indicates that the two profiles (one for the factors and subtests and one for the Language, Non-Language, and Total IQ's) may be studied from two points of view:

> Scores plotted above the 50th percentile and standard score line on the left-hand profile, or above the 100 IQ line on the right-hand profile, indicate areas in which the pupil is above average for his age level. Points below these two midpoints indicate areas in which he is below average. In this way, the individual is compared with others of his age.
>
> The zigzag nature of the plotted profile indicates how the pupil's performance on each component compares with his performance in other areas. This comparison facilitates an analysis of intraindividual differences.

In general, the seven test units are not interpreted separately but are combined, as in the profile, to produce three principal kinds of scores: factor, section, and total. An individual's unique pattern of mental functioning is said to be reflected by the factor and section scores, while the total provides an overall measure of the level of mental functioning.

According to the Manual, the abilities measured by the Language Section are generally regarded as more closely related to school-type tasks. The Non-Language Section measures abilities which are less dependent on verbal skills and is said to be less sensitive to socioeconomic and cultural influences.

References for Uses of the Test in Gifted Education

California Test Bureau. *Examiner's manual, California Short-Form Test of Mental Maturity*. Monterey CA: California Test Bureau, 1963.

CTB/McGraw-Hill. *Catalog*. Monterey CA: CTB/McGraw-Hill, 1974.

Tennessee Self Concept Scale: Counseling Form

Tennessee Self Concept Scale: Counseling Form by William H. Fitts. For individuals 12 years and older and having at least a sixth grade reading level. Published by Counselor Recordings and Tests, Box 6184, Acklen Station, Nashville TN 37212.

What the Test Measures

The Scale measures how the individual perceives himself in a variety of contexts.

Content and/or Structure of the Test

The Scale consists of 100 self administering items. Each item is answered using a 5 point scale: completely false; mostly false; partly false and partly true; mostly true; completely true. Ninety items are arranged in a 3×5 matrix system for measuring self concept. Items are arranged in the booklet in such a way that scores are obtained by counting row and column totals on the answer forms. The three rows represent items that measure Identity (What I am), Self Satisfaction (How I feel about myself), and Behavior (This is how I act). The five columns measure the different views of Self: Physical, Moral-Ethical, Personal, Family, and Social. The remaining 10 items yield a measure of self criticism which is used to help interpret the other scores.

How the Test Is Scored

Both manual and machine scoring of the Scale are possible. Instructions to the test taker request that no item be omitted. However, with all these instructions a respondent may still omit items. Scoring procedures require that omitted items be filled in by the scorer using the middle point for each (3—or partly false and partly true) before computing the score.

Hand scoring is facilitated by two types of answer forms. However, the Manual suggests computer scoring for 50 or more Scales. In view of the fact that 36 different subtotals and totals must be obtained from each sheet to get the 14 interpretable scores, it is suggested that scorers may wish to consider machine scoring at around 20 or 25.

Once scores are totaled they are plotted on a profile sheet for interpretation. To facilitate interpretation by the individual, two types of scores are provided, *T* scores and percentile scores. (*T* score refers to a standard score with a mean of 50 and a standard deviation of 10.)

What the Scores Mean

The Counseling Form has 15 interpretable scores: the Self Criticism Score; nine Positive Scores; three Variability Scores; the Distribution Score; and the Time Score.

The Self Criticism Score (SC)

The 10 items in this Scale were taken from the L-Scale of the Minnesota Multiphasic Personality Inventory. All 10 items are moderately derogatory statements that most people answer as being true for them. Individuals who answer most of these items false are considered as being defensive or they may be making a deliberate effort to present a favorable picture of themselves. High scores on the other hand generally indicate a normal, healthy openness.

Extremely high scores indicate that the individual may be lacking in normal defenses while extremely low scores indicate that the person is not being candid. The latter suggests that the Positive Scores may be somewhat elevated. Profiles of individuals with extremely high or low SC scores should be interpreted with caution.

The Positive Scores (P)

These scores represent a sort of internal frame of reference within which the individual describes himself. The statements fall into eight categories: This is what I am, This is how I feel about myself, This is what I do, Here is how I see my Physical, Moral-Ethical, Personal, Family, and Social Selves. Nine separate scores are obtained.

Total P Score. This is the most important single score on the Counseling Form. It reflects the overall level of self esteem. Generally, persons with high scores tend to like themselves, feel that they are of value and worth, have a great deal of self confidence, and tend to act accordingly. Persons with low scores tend to be doubtful about their own worth, see themselves as undesirable, often feel anxious, depressed, and unhappy, and have little self confidence.

Identity. These are the "What I am" items. These items attempt to describe how the individual sees himself.

Self Satisfaction. This set of items yields a score which reflects how the person feels about the self he perceives. In general this score reflects the level of self satisfaction or self acceptance.

Behavior. Here, the score comes from those items that say, "This is what I do, or, This is the way I act." This score then reflects the individual's perception of his own behavior.

Physical Self. This score presents the individual's view of his body, his state of health, physical appearance, skills, and sexuality.

Moral-Ethical Self. This score describes the self from the M-E frame of reference, that is, how one views himself in terms of moral worth, relationship to God, feelings of being "good" or "bad," et cetera.

Personal Self. This score gives an indication of the individual's sense of personal worth, feelings of adequacy as a person, and his evaluation of his personality apart from his body or relationship with others.

Family Self. This score indicates something of how the individual perceives himself in reference to his closest and most immediate circle of associates.

Social Self. This score indicates how the individual perceives himself in reference to "others" in a more general way. . . how the person sees himself and his worth in social interaction with other people in general.

The Variability Scores (V)

These scores give a measure of the variability or inconsistency from one area of self perception to another. High scores mean that the subject is quite variable in how he perceives himself in the various contexts. Low scores, in general, are indicative of a well integrated, self assured individual.

Total V. This represents the total amount of variability for the entire scale. High scores mean that a person's self concept is so variable from one area to another as to reflect little unity or integration. Well integrated individuals generally score below the mean but above the first percentile. Individuals scoring below the first percentile on Total V are considered "rigid."

Column Total V. Summarizes variations within five of the self concept areas: Physical, Moral-Ethical, Personal, Family, and Social Self.

Row Total V. Summarizes variations in the three internal frames of reference within which the individual is describing himself: Identity (What I am), Self Satisfaction (How I feel about myself), and Behavior (This is what I do).

The Distribution Score (D)

This score indicates the way the individual has distributed his responses in the five options (i.e., how many times were completely false, mostly false, partly false and partly true, mostly true, and completely true used?) High scores indicate that the individual is very definite and certain about what he says about himself. Low scores indicate the opposite. Low scores may also be indicative of defensiveness and guardedness. Extreme scores in either direction are considered undesirable and are often obtained from "disturbed" people.

The Time Score

This score is based on the total time (to the nearest minute) it takes the individual to complete the Scale. Since it has been found that individuals with sufficient education, intelligence, and reading ability are able to handle the task of completing the 100 questions in under 20 minutes, additional information about individuals taking over 20 minutes should be obtained. For example, other test scores or discussion with the individual to determine if other factors caused the excessive time (were they sick?, etc.) might clear up the problem.

References for Uses of the Test in Gifted Education

Fitts, W. H. *Tennessee Self Concept Scale: Bibliography of research studies.* (1st ed.) January 1973. (Over 300 references.)

Fitts, W. H. *Supplement to Tennessee Self Concept Scale: Bibliography of research studies.* February 1974. (Over 125 references.)

Lawrence, E. A., & Winschel, J. F. Self concept and the retarded: Research and issues. *Exceptional Children.* 1973. 39, 310-319.

McFern, A. R. A self concept study of adolescents in four areas of exceptionality. Unpublished doctoral dissertation, University of Georgia, 1973.

Self, P. A. Self concepts, attitudes, and values of women honor students. Unpublished doctoral dissertation, University of Texas at El Paso, 1973.

Torrance Tests of Creative Thinking

Torrance Tests of Creative Thinking by E. Paul Torrance. Appropriate for use in kindergarten through graduate school. Published by Personnel Press, Inc., Princeton NJ 08540.

What the Test Measures

Torrance defines creativity as "a process of becoming sensitive to problems, deficiencies, gaps in knowledge, missing elements, disharmonies, and so on: identifying the difficulty; searching for solutions, making guesses, or formulating hypotheses about the deficiencies; testing and retesting these hypotheses and possibly modifying and retesting them; and finally communicating the results" (Torrance, 1966, p. 6).
The test attempts to measure certain aspects of creativity. It does not attempt to measure all of the important dimensions of creative thinking. Four aspects of human behavior encompassed by Torrance's definition of creativity are sought in a verbal and a figural context: fluency, flexibility, originality, and elaboration.

Content and/or Structure of the Test

Verbal Activities

Ask-and-Guess. There are three activities here based on a drawing presented to the subject in the booklet. First, the individual is to write out all of the questions he can think of about the picture. Second, the test taker is to write as many possible causes of the action depicted in the drawing as he can. Finally, the examinee is to list as many possibilities of what might happen as a result of what is taking place in the picture. All three of these activities yield some measure of the four aspects of creativity—fluency, flexibility, originality, and elaboration—as defined.

Product Improvement. Here, the individual is to list clever, interesting, and unusual ways of changing a stuffed toy so children will have more fun playing with it. It permits individuals to play with ideas that would not normally be expressed in more serious tasks. The *fluency* score for this activity is the number of relevant responses produced. The *flexibility* score is the number of different approaches used in producing ideas for improvement. The *originality* score is based on the statistical infrequency and appropriateness of the ideas produced. The *elaboration* score is the number of different ideas or details given in elaborating or spelling out the ideas produced.

Unusual Uses. The subject is asked to list as many different and interesting things that he can think of for common, everyday items. Thus, this task is in part a test of the ability to free one's mind of well established sets. This activity yields scores for fluency, flexibility, originality, and elaboration determined in a manner similar to that described for the Ask-and-Guess activity.

Unusual Questions. In this activity the subject is to think of as many questions as he can about the everyday item for which he described an Unusual Use in the previous activity. The directions suggest that the individual think of questions which would lead to a variety of different answers and might arouse interest and curiosity in others. This activity yields scores on all four measures of creative thinking.

Just Suppose. This task is an adaptation of the consequences type test used in Guilford's battery. The subject is confronted with an improbable situation and asked to predict the possible outcomes from the introduction of a new or unknown variable. This activity purports to measure the subject's ability to consider, evaluate, and play with unusual ideas and to think through their possible consequences. This measures all four aspects of creative thinking.

Figural Tasks

Picture Construction. This activity requires the subject to think of a picture in which a given shape made of colored paper is an integral part. Elaboration is encouraged by the instructions to add ideas that will make the picture tell as complete and as interesting a story as possible. The product is evaluated only for originality and elaboration.

Incomplete Figures. In this task, a number of incomplete figures are presented and the subject is to produce an original response to complete the figures. Each figure completed is scored for flexibility, originality, and elaboration. Titles are scored for originality and cleverness. Since about one-third of the subjects complete all 10 figures, this activity provides a fluency score of only moderate usefulness.

Repeated Figures. The stimulus material here is either 30 parallel lines or 40 circles. The task requires the ability to make multiple associations to a single stimulus. A deliberate attempt is made

here to stimulate all four types of divergent thinking.

How the Test Is Scored

Each of the seven verbal tasks and each of the three pictorial tasks have explicit directions for scoring in a Directions Manual and Scoring Guide. Each Scoring Guide emphasizes that it is most important for the scorer to read and follow the guide as precisely as possible, accepting the standards of the guide as a basis for judgment.

First, and most important according to the Guide, the scorer must have an understanding of the concepts of fluency, flexibility, originality, and elaboration. If such an understanding does not exist, it is suggested that the introductory sections of the Norms-Technical Manual be read or reread. Second, the scorer should be familiar with the rationale for the seven verbal and three pictorial test tasks.

In a preliminary way, the scorer must determine whether a response is scoreable, that is, has relevance to the test activity. Next, he rereads the scoring guide with a completed record, finding the responses on it in the lists of scoring categories. No entries are made on worksheets for responses determined "not scoreable."

For the verbal tasks, compound sentences containing two or more distinct ideas should be treated as two or more responses. However, do not confuse "single idea" sentences containing elaborations or details with multiple responses. Now scoring can begin. An example of a scoring worksheet is contained in the Scoring Guide. After entering the desired identifying information proceed with the scoring of each activity as follows:

Scoring Ask Questions. Determine the originality weight and flexibility category numbers and record them in the appropriate boxes under Activity 1 for each response. If the response does not appear on the list given, it is a rather infrequent one. If it involves creative strength and gets away from the obvious, it should be given a weight of 2. Creative strength is described as requiring more intellectual energy than giving obvious, common, and learned responses. Scorers are advised to have the concept of creative strength well in mind as they score protocols.

Scoring Guess Causes. Follow the steps outlined for Activity 1, remembering that a response is regarded as "inadequate" and is not scoreable if causality cannot be inferred.

Scoring Guess Consequences. Follow the same procedure, remembering that an adequate response must involve some inferable effect relationship.

Scoring Product Improvements. Same as preceding activities. Scoreable responses here must retain some quality of a toy and "be fun to play with."

Scoring Unusual Uses. Same as above. A scoreable response here must involve some use of a cardboard box or tin can, either intact or disassembled and modified.

Scoring Unusual Questions. Fluency is scored as in all others, but no score is determined for flexibility and elaboration. Using the definitions of "divergent power" and the examples given in the Scoring Guide, determine the originality score for this activity.

Scoring Just Suppose. In this activity, the flexibility score is defined as a "change or shift in attitude or focus." The first response is not scored, and if the attitude or focus in all responses is similar, the flexibility score is zero. If changes in attitude or focus do occur, each receives a point. Once a shift has occurred, duplications do not receive additional credit. Originality is determined by checking each response against the list of zero and one credit responses in the guide. Responses not included in the guide which get away from the very obvious receive 2 credits.

Summarizing results for the Verbal Activities. After each of the verbal activities has been scored, a Score Summary should be computed. The fluency score is simply the total derived from each activity. To determine the flexibility score in Activities 1-5, strike out category duplications, count the remaining responses, and add to the total number of checkmarks under the "Shift" column for Activity 7. There is no flexibility score for Activity 6. The Originality score is obtained by adding the weights recorded in the originality column for each activity. No provision has been made for determining a verbal elaboration score.

Scoring Picture Construction. Using the Scoring Guide, determine the originality weight for the response and place this score in the box labeled originality. Next figure the elaboration score and place this score in the appropriate box. Since there is only one response for this activity, it is not scored for flexibility or fluency.

Scoring Picture Completion. Again, using the Scoring Guide, determine the originality weight and flexibility category for the first response and enter these in the appropriate places. Determine the elaboration score of the first response and

enter it in the elaboration column. Continue in this fashion for each scoreable response.

Scoring Parallel Lines. Using the Guide, determine for each scoreable response the flexibility category, the originality weight, and the elaboration score and record in the appropriate boxes of the scoring sheet.

Summarizing results for the Pictorial Activities. No flexibility or fluency scores are reported for Picture Construction. Simply transfer the originality and elaboration scores to the appropriate places on the summary sheet. For Picture Completion, count the number of scoreable responses as the fluency score. The flexibility score is the number of responses after striking out duplications. Originality and elaboration scores are obtained by totaling the weights assigned during the scoring phase. Parallel Lines is scored the same as Picture Completion, though bonus points for originality may be awarded for combining two or more figures. The procedures are outlined in the Scoring Guide.

What the Score(s) Mean(s)

Cautions

Individuals may manifest creative abilities in an almost infinite variety of ways. Therefore, be aware that only selected abilities are being assessed by these activities, not the entire range of creative abilities.

Second, the user is warned that these abilities do not follow the same developmental pattern that is typically found for intelligence and achievement. Most users have found sharp drops in performance usually at the beginning of Grades 4 and 7. Recent studies, however, have shown some reversal in this developmental pattern.

Users of this set of exercises are encouraged to base their interpretations on the three verbal and the four figural scores rather than to rely on a single score in isolation. A single score may be misleading if not interpreted in light of the other scores. For example, an extremely high fluency score would be interpreted quite differently if accompanied by either very high or very low originality and flexibility scores. Many people could produce a very large number of common, obvious, trite responses which would produce a very high fluency score. If the originality and flexibility scores accompanying this fluency score are very low, one would not interpret this as a brilliant display of creative energy.

Verbal Scores

Fluency. This score reflects the ability to produce a large number of ideas with words. Since there are seven activities or tasks which contribute to this score, each of which taps a slightly different kind of ability or mental process, one should look at the subject's productions across the separate tasks.

Flexibility. This score reflects the person's ability to produce a number of kinds of ideas, to go from one approach to another, or to use a variety of strategies. At times it may be helpful to interpret this score in relation to verbal fluency.

Originality. This score represents the ability to produce ideas that are different from the obvious, commonplace, trite, or well established. Viewed in relation with the Verbal Flexibility score, this score can provide a measure of the "degree of divergency" in thinking.

Elaboration. Though no norms have been established for this score and interscorer reliability of untrained scorers is low, some users find this score useful. High scores seem to be related to school achievement and usually characterize persons who have incentive and take constructive action. Low scores usually characterize school dropouts, delinquents, and underachievers.

Figural Scores

Fluency. This score is very useful in helping the user understand the other figural scores. Impulsive thinkers and even "non-thinkers" can achieve high scores on this activity rather easily. Such persons however usually always have low scores on the other three figural activities. Low scores may occur because the subject does a great deal of elaborating or if the subject blocks, spends a great deal of energy trying to get away from the commonplace, or is not well motivated.

Flexibility. Interpreting this score is basically the same as for Verbal Flexibility except that it is concerned with figural rather than verbal modes of thinking.

Originality. Anchors to interpretation can be derived by looking at this score in relation to the Fluency and Elaboration scores.

Elaboration. This score measures the subject's ability to develop, embellish, carry out, or otherwise elaborate ideas.

Title Originality

No comparison group has been supplied and the score for the activity is not included in the composite verbal scores. It can be useful since it gives the test taker an opportunity to show how he is "warmed up" by his own productive thinking.

Low scores seem to be characteristic of underachievers, school dropouts, delinquents, and extreme conformists.

Verbal-Figural Discrepancies

Large differences—ten or more T score points—may be useful in understanding a person's mental functioning, assessing potentialities, or guiding them in achieving more healthy development. Caution should be exercised in making interpretations of the relationship between these two scores and suggesting courses of action on the basis of observed discrepancies.

Change Scores

Because the author and publisher have encouraged use of the tests as research instruments, obviously they could and should be used in pre- and posttest designs in the study of creativity and in longitudinal studies of creativity.

References for Uses of the Test in Gifted Education

Even, R. L. An experimental study of the comparative effect of selected art experiences on the creative performance and attitudes of academically superior students. Doctoral thesis, University of Minnesota, 1963. D.A. 24:4470.

Fleming, E. S., & Weintraub, S. Attitudinal rigidity as a measure of creativity in gifted children. *Journal of Educational Psychology*, 1962, 53, 81-85.

Torrance, E. P. *Gifted children in the classroom*. New York: Macmillan, 1965.

Torrance, E. P. *Torrance Tests of Creative Thinking. Directions manual and scoring guide, Verbal Test, Booklets A & B* (research ed.). Princeton NJ: Personnel Press, 1968.

Torrance, E. P. *Torrance Tests of Creative Thinking. Directions manual and scoring guide, Figural Test, Booklets A & B* (research ed.). Princeton NJ: Personnel Press, 1970.

Torrance, E. P. *Torrance Tests of Creative Thinking. Norms-technical manual, Verbal Tests, Forms A & B, Figural Tests, Forms A & B* (research ed.). Princeton NJ: Personnel Press, 1966.

Weisberg, P. S., & Springer, K. J. Environmental factors in creative functions. *Archives of General Psychiatry*, 1961, 5, 554-564.

Scales for Rating Behavioral Characteristics of Superior Students

Scales for Rating Behavioral Characteristics of Superior Students by Joseph S. Renzulli, Linda H. Smith, Alan J. White, Carolyn M. Callahan, and Robert K. Hartman. Grades or ages for which appropriate not specified. Published by Joseph S. Renzulli, Bureau of Educational Research, School of Education, University of Connecticut, Storrs CT 06268.

What the Test Measures

The 10 Scales represent "an attempt to provide a more objective and systematic instrument that can be used as an aid to guiding teacher judgment in the identification [of gifted] process" (Renzulli, et al., n.d., p. 1). According to the authors, the Scales go beyond traditional rating scales which typically included only measures of learning and motivation. These scales also tap creativity, leadership, artistic, musical, dramatic, communication (precision and expressiveness), and planning characteristics.

Thus, the Scales attempt to tap teacher judgment in identifying those students who might be classified as gifted or talented according to both the traditional definition of "intellectually gifted" and the more divergent definition of gifted which includes such attributes as creativity, leadership, and motivational, artistic, musical, and dramatic talents (Renzulli, et al., n.d., pp. 4-11).

Content and/or Structure of the Test

Each of the Scales requires the teacher or other person completing the form to respond to items presented by checking one of four options: seldom or never, occasionally, considerably, or almost always. The descriptions that follow were taken from the Manual and Scales (Renzulli, et al., n.d., pp. 38-48).

Learning Characteristics

This scale requires the rater to react to eight items related to the student's learning "behavior." Items ask the rater to judge quality of the pupil's vocabulary; general information level; recall of factual information; insight into cause-effect relationships; how well the pupil grasps basic principles and makes generalizations about people, things, et cetera; how well the pupil "observes" (i.e., "sees more" or "gets more" out of stories, films, etc.); type and quality of reading undertaken; and how well complicated material is understood.

Motivational Characteristics

This scale consists of nine items. Raters are asked to judge whether or not the pupil in question is truly involved and persistent in seeking task completion, is easily bored with routine tasks, needs little external motivation, strives toward perfection, prefers to work independently, is interested in many "adult" problems, is often self assertive, likes to organize and structure, and is quite concerned with right and wrong, good and bad, et cetera.

Creativity Characteristics

This 10 item rating scale asks the rater to judge the degree to which the pupil displays curiosity; generates ideas which are often unique; is uninhibited in expressing opinions; is a high risk taker; displays a good deal of "intellectual playfulness," fantasy, and imagination; displays a keen sense of humor; is unusually aware of impulses; is sensitive to beauty or the aesthetic characteristics of things; is nonconforming; and critizes constructively.

Leadership Characteristics

Pupils are rated here according to the degree to which they carry responsibility, are self confident with others, are liked by others, cooperate with teacher and peers, express themselves, adapt to new situations, enjoy being around other people, tend to dominate others by directing activities, participate in most social activities, and excel in athletic activities. There are 10 items to be rated in this scale.

Artistic Characteristics

This scale has 11 items. Pupils are rated as to the degree to which they like to participate in art activities, incorporate a number of elements into art work, arrive at unique solutions to artistic problems, concentrate for long periods on art projects, try out different media, select art media for free activities, are sensitive to the environment, produce balance and order in art work, are critical of their own work, show interest in other students' art work, and elaborate on ideas from other people.

Musical Characteristics

Pupils are rated on how often they show a sustained interest in music, perceive fine differences in musical tone, easily remember melodies, participate in musical activities, play a musical instrument or indicate a desire to, are sensitive to the rhythm of music, and are aware of and can identify a variety of sounds heard at a given moment. Seven attributes are rated in this scale.

Dramatic Characteristics

Ten attributes are rated in this scale in terms of how often pupils volunteer to participate in classroom plays or skits, tell stories or give an account of some experience, use gestures and facial expressions to communicate feelings, role play or act out situations, can readily identify themselves with the moods and motivations of characters, handle body with ease and poise, create original plays or make up plays from stories, hold the attention of a group when speaking, can evoke emotional responses from listeners, and can imitate others.

Communication Characteristics

Precision. Eleven characteristics are rated in terms of how often an individual speaks and writes directly to the point, modifies and adjusts expression of ideas for better reception, reviews and edits in a way which is concise, explains or expresses things or thoughts clearly and precisely, uses descriptive words to add color or emotion, can find a number of ways to express ideas so others will understand, describes things in a few appropriate words, expresses fine shades of meaning by use of a large stock of synonyms, expresses ideas in a variety of alternate ways and knows and can use many words closely related in meaning.

Expressiveness. Pupils are rated on four attributes: their use of voice expressively to convey or enhance meaning, their use of nonverbal expression to communicate more effectively, their effectiveness as storytellers, and their use of colorful and imaginative figures of speech such as puns or analogies.

Planning Characteristics

Fifteen items or attributes are rated in this scale. Each is rated in terms of how often a pupil determines what information or resources are needed to complete a task, understands the relationship of the parts to the whole, allows time to complete an entire task, foresees consequences or effects of actions, organizes work, takes all details into account which are necessary to complete a task, demonstrates good strategy where necessary, recognizes alternative methods for accomplishing a goal, can identify or anticipate areas of difficulty, arranges steps to be taken into a sensible order or time sequence, can break down an activity into step by step procedures, establishes priorities, understands limitations relating to various aspects of the project, can provide details that contribute to the development of a plan or procedure, and can see alternative ways to distribute work or assign people to accomplish a task.

How the Test Is Scored

Separate scores for each of the 10 dimensions are obtained as follows:

Add the total number of checks or X's in each column to obtain the Column Total.

Multiply the Column Total by the Weight for each column to obtain the Weighted Column Total. (Note that in all cases Seldom or Never receives a weight of 1, Occasionally receives a weight of 2, Considerably receives a weight of 3, and Almost Always receives a weight of 4.)

Sum the Weighted Column Totals across to obtain the "Score" for each dimension of the scale.

What the Score(s) Mean(s)

Though the Manual does not specifically state, one infers from the weighting of the option Almost Always that this is the desired rating for those "who exhibit gifted behavior." Thus, high scores on any of the scales should be interpreted as exhibiting "gifted or talented" behavior even though no distributions of scores, norms, or other interpretive aids are presented. Maximum scores for each of the dimensions are not presented though they can be easily determined by the user.

The authors do offer general comments concerning the use and interpretation of the scales. The first comment is that only those scales that are relevant to program objectives should be selected for use in a particular program. In other words, some selectivity should be exercised in deciding which of the scales to use. Not only are some not appropriate for some programs, but as the Manual points out, asking teachers to complete all of the scales for a group of students is a burdensome task and could result in superficial or hastily completed ratings.

The Manual also points out that the first three scales—learning, motivation, and creativity—are consistent with the objectives of most programs for the gifted and talented. The authors do

not say so, but one infers, that these three should be used in most situations, while the remaining seven should be selectively chosen depending on the program objectives.

Because the Scales represent 10 diverse dimensions of behavioral characteristics related to gifted or talented behavior, the authors caution against adding subscores (dimensions) together to form a total score.

Though the authors indicate that no predetermined cutoff scores for the scales can be provided, they suggest that users compute a mean score on each dimension for the total number of students being considered for enrollment in a special program and that those students who are "significantly" above the mean be selected for placement in that program or activity.

References for Uses of the Test in Gifted Education

Renzulli, J. S., et al. *Scales for Rating Behavioral Characteristics of Superior Students, Examiner's Manual*. Storrs: Bureau of Educational Research, School of Education, University of Connecticut, n.d.

Renzulli, J. S., Hartman, R. K., & Callahan, C. M. Teacher identification of superior students. *Exceptional Children*, 1971, *38*, 211-214.